Course Title: Data Analysis Course Syllabus for Beginners and Intermediate

Course Description: This course introduces data analysis for beginners and intermediate learners. Students will learn fundamental concepts, techniques, and tools for collecting, cleaning, exploring, and visualizing data. The course covers both theoretical foundations and practical applications of data analysis using Python and popular libraries such as pandas, NumPy, Matplotlib, Power BI, MySQL.

Prerequisites: Basic knowledge of programming concepts and familiarity with Python programming language. No prior experience in data analysis is required.

Course Objectives:

1. Understand the fundamentals of data analysis and its importance in decision-making processes.
2. Learn how to collect, clean, and preprocess data for analysis.
3. Gain proficiency in using Python libraries such as pandas and NumPy for data manipulation and analysis.
4. Develop skills in data visualization using Matplotlib and other visualization libraries.
5. Explore exploratory data analysis (EDA) techniques and statistical methods for deriving insights from data.

Course Outline:

Module 1: Introduction to Data Analysis

- Overview of data analysis and its applications
- Introduction to Python for data analysis
- Setting up Python environment and Jupyter Notebook

Module 2: Data Collection and Cleaning

- Data types and formats
- Data collection methods (web scraping, APIs, databases)
- Data cleaning techniques (handling missing values, duplicates, outliers)

Module 3: Introduction to Pandas

- Introduction to panda's library
- Series and DataFrame objects
- Data manipulation and transformation with pandas

Module 4: Data Visualization with Matplotlib

- Introduction to Matplotlib library
- Basic plots (line plots, bar plots, scatter plots)
- Customizing plots and adding annotations

Module 5: Exploratory Data Analysis (EDA)

- Understanding the structure and distribution of data
- Summary statistics and descriptive analysis
- Correlation analysis and heatmap visualization

Module 6: Data Analysis with NumPy

- Introduction to NumPy library
- Array creation and manipulation
- Mathematical operations and statistical functions

Module 7: Advanced Data Analysis Techniques

- Grouping and aggregation with pandas
- Time series analysis and visualization
- Introduction to machine learning for data analysis

Module 8: Data Visualization with Seaborn

- Introduction to Seaborn library for statistical data visualization
- Advanced plots (box plots, violin plots, pair plots)
- Visualizing relationships and patterns in data

Module 9: Introduction to Statistical Analysis

- Basic concepts of statistical analysis
- Hypothesis testing and p-values
- Regression analysis and model fitting

Module 10: Real-world Data Analysis Projects

- Working on real-world datasets
- Applying data analysis techniques learned throughout the course
- Presenting findings and insights from data analysis projects

Module 11: Intermediate Data Analysis

- Advanced data cleaning techniques
- Handling large datasets with pandas and Dask
- Advanced statistical analysis and machine learning concepts

Module 12: Introduction of MySQL Database Server

❖ **Introduction to MySQL**
  ➢ Overview of relational databases
  ➢ Introduction to MySQL RDBMS
  ➢ Installing and setting up MySQL server

❖ **MySQL Basics**
  ➢ MySQL data types
  ➢ Creating and managing databases
  ➢ Creating tables and defining constraints

❖ **SQL Basics**
  ➢ Introduction to SQL (Structured Query Language)
  ➢ SELECT statement and querying data
  ➢ Filtering, sorting, and limiting results

❖ **Data Manipulation with SQL**
  ➢ Inserting, updating, and deleting data
  ➢ Modifying table structure
  ➢ Working with NULL values

❖ **Advanced SQL Queries**
  ➢ Joins (inner, outer, self)
  ➢ Subqueries
  ➢ Aggregate functions and grouping

❖ **Indexes and Performance Optimization**
  ➢ Understanding indexes and their types
  ➢ Indexing strategies for performance optimization
  ➢ Analyzing and optimizing queries

❖ **Transactions and Concurrency Control**
  ➢ Introduction to transactions
  ➢ ACID properties
  ➢ Locking mechanisms and concurrency control

❖ **Stored Procedures and Functions**
  ➢ Creating and executing stored procedures
  ➢ Creating and using user-defined functions
  ➢ Benefits and best practices

❖ **MySQL Administration**
  ➢ User account management

- ➢ Backup and restore operations
- ➢ Security considerations
- ❖ **Database Design Best Practices**
- ➢ Normalization and denormalization
- ➢ Designing efficient database schemas
- ➢ Data integrity and constraints

13. <mark>Introduction to Power BI:</mark>

- ➢ Overview of Power BI and its components.
- ➢ Understanding the Power BI ecosystem: Power BI Desktop, Power BI Service, and Power BI Mobile.

❖ Data Preparation and Import:

- ➢ Connecting to various data sources: Excel, databases (SQL Server, MySQL, etc.), online services (Google Analytics, Salesforce, etc.), and files (CSV, JSON, etc.).
- ➢ Data transformation using Power Query Editor: cleaning, shaping, merging, and appending data.

❖ Data Modeling:

- ➢ Understanding relationships: one-to-one, one-to-many, and many-to-many.
- ➢ Creating and managing relationships between tables.
- ➢ Implementing data modeling best practices for efficient analysis.

❖ Data Visualization:

- ➢ Creating interactive reports and dashboards.
- ➢ Choosing appropriate visualizations: charts (bar, line, pie, etc.), maps, tables, and matrices.
- ➢ Formatting and customizing visuals to improve aesthetics and readability.

❖ Advanced Analytics with DAX:

- ➢ Introduction to Data Analysis Expressions (DAX).
- ➢ Writing DAX formulas for calculated columns, measures, and calculated tables.
- ➢ Performing advanced calculations and aggregations using DAX functions (SUMX, CALCULATE, etc.).

❖ Data Exploration and Insights:

- ➢ Utilizing features like drill-down, drill-through, and cross-filtering for interactive exploration.
- ➢ Creating hierarchies and implementing filtering options for deeper analysis.
- ➢ Leveraging AI-powered insights and data-driven storytelling.

❖ Data Sharing and Collaboration:

- ➢ Publishing reports and dashboards to the Power BI Service.
- ➢ Configuring sharing settings and permissions for collaboration.
- ➢ Embedding reports in other applications or websites.

❖ Data Refresh and Scheduled Refresh:

- ➤ Setting up data refresh schedules to keep reports up-to-date.
- ➤ Configuring data source credentials for automated refresh.
- ➤ Troubleshooting common data refresh issues.

❖ **Advanced Power BI Features:**

- ➤ Implementing row-level security to restrict data access based on user roles.
- ➤ Creating custom visuals using Power BI Custom Visuals SDK.
- ➤ Implementing complex data models and calculations for specialized analytical needs.

❖ **Performance Optimization and Best Practices:**

- ➤ Optimizing report performance for faster loading and rendering.
- ➤ Applying best practices for report design, data modeling, and visualization.
- ➤ Monitoring and optimizing Power BI Service performance for efficient usage.

14.   Introduction to Advanced Excel Features:

- ➤ Review of basic Excel functions and formulas.
- ➤ Introduction to advanced features such as PivotTables, Power Query, and Power Pivot.

❖ **Data Cleaning and Preparation:**

- ➤ Techniques for cleaning and transforming raw data.
- ➤ Removing duplicates, handling missing values, and formatting data for analysis.

❖ **Data Analysis Tools:**

- ➤ Advanced filtering and sorting techniques.
- ➤ Goal Seek and Solver for optimization problems.
- ➤ Scenario Manager for sensitivity analysis.

❖ **Statistical Analysis in Excel:**

- ➤ Descriptive statistics: mean, median, mode, standard deviation, etc.
- ➤ Correlation analysis: calculating correlation coefficients.
- ➤ Regression analysis: performing linear regression analysis to model relationships between variables.

❖ **Advanced Functions and Formulas:**

- ➤ Array formulas for performing calculations on arrays of data.
- ➤ Lookup and reference functions: INDEX-MATCH, VLOOKUP, HLOOKUP, etc.
- ➤ Logical functions: IF statements, nested IFs, AND, OR, etc.

❖ **Visualizing Data with Charts and Graphs:**

- ➤ Creating advanced charts like waterfall charts, Gantt charts, and box plots.
- ➤ Customizing chart elements and formatting options.
- ➤ Using Sparklines for visualizing trends within cells.

❖ **Data Visualization with PivotTables and PivotCharts:**

- ➤ Creating PivotTables to summarize and analyze data.
- ➤ Creating PivotCharts to visualize PivotTable data.

➢ Slicers and Timelines for interactive data analysis.

❖ **Data Modeling with Power Pivot:**
➢ Introduction to Power Pivot and its advantages for data analysis.
➢ Importing data from multiple sources into Power Pivot.
➢ Creating relationships between tables and DAX formulas for advanced calculations.

❖ **Automating Tasks with Macros and VBA:**
➢ Recording and editing macros to automate repetitive tasks.
➢ Introduction to Visual Basic for Applications (VBA) programming.
➢ Writing custom VBA macros for advanced automation and data manipulation.

❖ **Data Analysis Case Studies:**
➢ Applying advanced Excel techniques to real-world data analysis scenarios.
➢ Solving complex business problems using Excel's analytical capabilities.

Module 15: Capstone Project

- Developing a comprehensive data analysis project
- Identifying a research question or problem statement
- Collecting, cleaning, analyzing, and visualizing data to derive insights

Assessment:

- Weekly assignments to reinforce learning concepts.
- Midterm project: Performing exploratory data analysis on a given dataset.
- Final project: Developing a complete data analysis project from data collection to visualization and presenting findings.

Textbook: "Python for Data Analysis" by Wes McKinney

Additional Resources:

- Online tutorials and documentation (pandas documentation, NumPy documentation, Matplotlib documentation, etc.).
- Supplemental readings and materials provided by the instructor.

Grading:

- Assignments: 30%
- Midterm Project: 20%
- Final Project: 40%
- Participation and Attendance: 10%

CSDT Centre Patna

Attendance Policy: Regular attendance is expected. Students are allowed a maximum of three unexcused absences. Excessive absences may result in a reduction of the final grade.

Office Hours: Instructor office hours will be held twice a week for additional help and clarification.